



*Citation for published version:*

Jones, K 2011, *Assessing institutional data storage and management using the Data Asset Framework (DAF) methodology at the University of Bath*. University of Bath, Bath.

*Publication date:*  
2011

[Link to publication](#)

**University of Bath**

### **Alternative formats**

If you require this document in an alternative format, please contact:  
[openaccess@bath.ac.uk](mailto:openaccess@bath.ac.uk)

#### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

#### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



## **DAF Report**

### **Assessing institutional data storage and management using the Data Asset Framework (DAF) methodology**

**University of Bath**

**May 2011**

Compiled by Kara Jones

Data gathered by Katie Evans, Jez Cope and Lizzie Richmond.

With thanks to those who completed the online survey, and those who generously provided their time and experience for interviews.

**Summary:**

The JISC funded Data Asset Framework (DAF) methodology was used to gather information and evidence on current research data management practices around the University of Bath. Details were gathered using an online survey (54 responses) and 23 face-to-face interviews with a variety of researcher-active staff and postgraduate students.

Several themes emerged from the survey and interviews:

1. Most respondents are satisfied with their data storage, although insufficient storage space is the biggest data management issue. There is an over-reliance on external hard-drives for backing-up data. This is a concern for the retrievability and security of data;
2. Data management practices are guided by intuition rather than informed by good practice;
3. There is a need for services to share research data during the life of the project with internal and external partners. These should be easy to use, geographically accessible, and employ version control;
4. Data reuse and sharing is considered a good idea but very few researchers are currently actively sharing their data, or providing robust metadata for archival storage;
5. Guidance on, and storage for the preservation of data would be appreciated.

Data is stored, shared and used differently at different parts of the research lifecycle. Different storage space and access requirements exist particularly during the working life of the data, and then during the period in which the data is archived.

Most of the data gathered provides evidence to support anecdotal material on current data management practices. Solutions to these issues include advocacy and training for data management plans, and robust storage with mechanisms for sharing throughout the lifecycle of the data.

Research data storage activities could be piloted by the following groups whose members noted different practices and requirements during the interviews: The Centre for Pain Research, the Doctoral Training Centre (Chemistry), Mechanical Engineering (Aerospace), Humanities & Social Sciences Graduate School, a school of Management centre and the Centre for Electron Optical Studies (as an analytical service).

## 1. **Rationale:**

This audit was undertaken to provide evidence of current research data practices through the University. This information has been collected for the Research Data Management Steering Group, sponsored by the Pro-Vice-Chancellor (Research), and will be used to inform recommendations on University training, policy and service provision for research data.

### **Aims:**

1. To investigate the types of data held by researchers throughout the university, researchers' existing data management practices and any issues or risks associated with these practices.
2. To provide evidence to inform University data management policy and services to satisfy the requirements of researchers and funders.

## 2. **Methodology:**

Information on research data practices were gathered using the JISC funded Data Asset Framework (DAF) with modifications for local issues.

The DAF is a methodology for assessing data management and holdings in an institution. Initially known as the Data Audit Framework, the development of DAF was funded by the Joint Information Systems Committee (JISC) following a recommendation from the report 'Dealing with Data'<sup>1</sup>. Four stages are set out in the DAF Implementation Guide<sup>2</sup> as follows:

**Stage 1** is for planning, defining the purpose and scope of the survey and conducting preliminary research.

**Stage 2** is about identifying what data assets exist and classifying them to determine where to focus efforts for more in-depth analysis.

**Stage 3** is where the information life cycle is considered to understand researchers' workflows and identify weaknesses in data creation and curation practices.

**Stage 4** pulls together the information collected and provides recommendations for improving data management.

---

<sup>1</sup> Lyon, L. (2007) *Dealing with data: Roles, rights, responsibilities and relationships* [online]. Available from: <http://www.jisc.ac.uk/whatwedo/programmes/digitalrepositories2005/dealingwithdata.aspx> [Accessed 4 April 2011].

<sup>2</sup> Digital Curation Centre (2009) *Data Asset Framework: Implementation guide* [online]. Available from: [http://www.data-audit.eu/docs/DAF\\_Implementation\\_Guide.pdf](http://www.data-audit.eu/docs/DAF_Implementation_Guide.pdf) [Accessed 4 April 2011].

### 3. Implementation

#### Stage 1: Planning the audit

The DAF was used to underpin a project run within the Research Data Management Steering Groups' remit. A project proposal was introduced to the group on 25 January 2011. Revisions after discussions with group members were made, with an undertaking to run the project during March-April 2011, reporting back by May 2011 (Appendix 1).

The Research Publications Librarian acted as auditor, with initial background reading and research on the framework implementation and methodology. The *Audit Form 1* in Appendix 2 gives an outline of the institution and some existing related documentation. The DAF project at the University of Northampton<sup>3</sup> provided a useful guide for running the activity.

#### Stages 2 and 3: Identifying and classifying data assets and Assessing the management of data assets

The Data Audit Framework Methodology<sup>4</sup> guide was used to inform the project, however the level of detail was far more involved than the set time frame, so a DAF 'lite' approach was taken using the toolkit. For example, we did not undertake a comprehensive analysis of documentary sources in departments but rather gathered details using the *Audit Form 3A: Data Assessment Management (Core element set)* for one data set from each interviewee.

A survey was prepared using the *Bristol Online Survey* tool. At the request of the steering group, the survey was anonymous to encourage responses that were honest and open. The survey was advertised and disseminated via a variety of channels, gaining 54 responses. This low number of results may have been the result of 'survey fatigue' with a number of surveys operational at that time.

A more detailed response to the project came from a follow-on series of interviews with a variety of researchers across the university. Interviews were conducted by Kara Jones (Library), Katie Evans (Library), Lizzie Richmond (Library), and Jez Cope (DTC Chemistry). The interviews used a modified version of the templates provided in the Data Asset Framework Implementation Guide<sup>5</sup>, particularly that of the University of Southampton<sup>6</sup>.

By Faculty/School, interviewees came from: Engineering (7), Science (6), H&SS (7), Management (3). By researcher role, interviewees fell roughly into categories of: PhD (6), Research Officer/PostDoc (6), Principle Investigator/Project Manager (11).

---

<sup>3</sup> Picton, M, Alexogiannopoulos, E & S. McKenney (2010). *Research Data Management Project: A DAF investigation of research data management practices at The University of Northampton. Final Report*. Northampton: University of Northampton. Available: <http://nectar.northampton.ac.uk/2736>. Accessed 5 March 2011.

<sup>4</sup> Jones, S, Ross, S & R Ruusalepp (2009). *Data Audit Framework Methodology, draft for discussion, version 1.8*. Glasgow: HATII, University of Glasgow.

<sup>5</sup> Data Asset Framework Implementation Guide. Available: [http://www.data-audit.eu/docs/DAF\\_Implementation\\_Guide.pdf](http://www.data-audit.eu/docs/DAF_Implementation_Guide.pdf)

<sup>6</sup> McGowan, T. & Gibbs, T. A. (2009) Southampton Data Survey: Our Experiences & Lessons Learned [unpublished]. University of Southampton: UK. Available: [http://www.data-audit.eu/docs/DAF\\_Implementation\\_Guide.pdf](http://www.data-audit.eu/docs/DAF_Implementation_Guide.pdf), pp. 25-31.

## Stage 4: Reporting results and making recommendations

This report is the outcome of the previous stages. It collates and analyses information from the audit, identifies weaknesses in data management and provides recommendations for improvement.

### 4. Findings:

Information gathered via the survey and the interviews is reported here. The survey was structured in 5 sections with responses provided below. Further evidence from interviews has been recorded where appropriate.

Survey sections are:

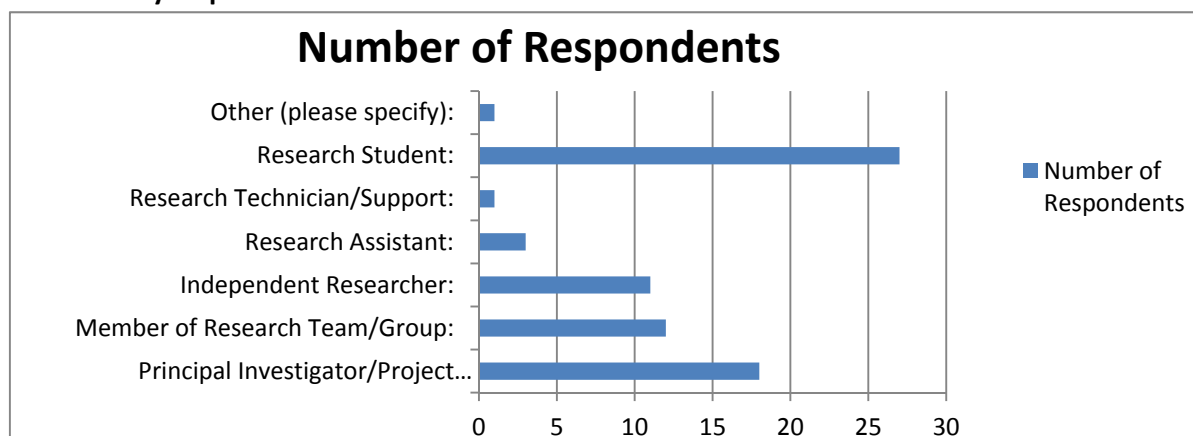
1. Characteristics of data produced by our researchers
2. Storage and backup activities
3. Access and data sharing
4. Data management
5. Data preservation

Respondents were able to choose more than one response for some questions.

#### 4.1 Characteristics of data produced by our researchers

In this section we asked respondents to describe their role, and the types of data they currently produced. We wanted a brief idea of whether respondents were documenting their data and who they thought owned the material. The survey responses are matched by the interview responses in terms of the types of data produced or collected, and the perceived ownership of the data.

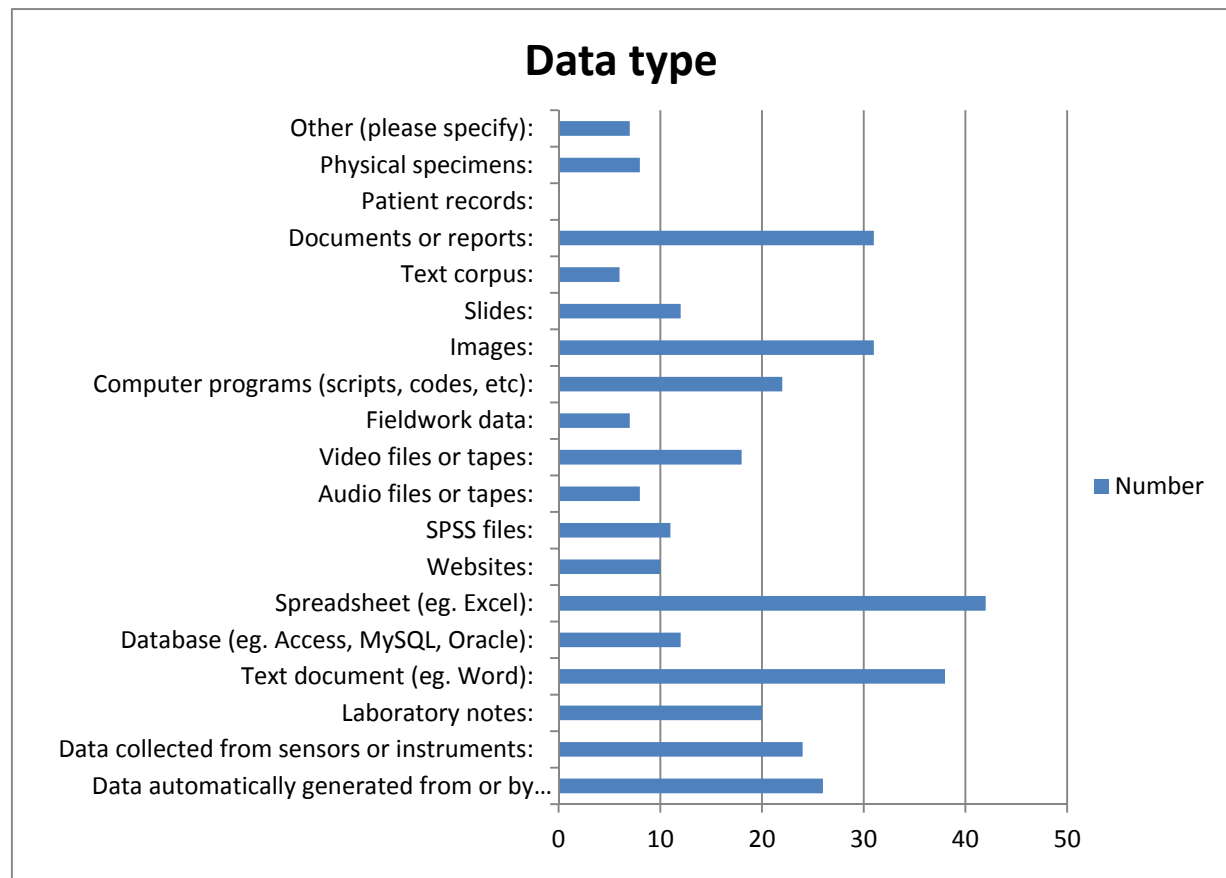
##### 4.1.1 Survey respondents



The online survey attracted a total of 54 respondents, largely research students and Principal Investigators/Project Managers. Information about the respondents department was not gathered. Interviews gathered further details from across the departments and academic roles. Responses from these have been anonymised where requested by the interviewee.

#### 4.1.2 Types of data being collected/needing support.

Data is a term with broad-ranging, diverse characteristics, often depending on discipline. Some of the main format types were suggested as responses for this question.



Multiple responses were permitted in this question. Some fields overlap, however a large number of respondents produce word processing, or spreadsheet file types, and a significant portion of respondents use image files, or data generated or collected from machines. The large number of image and machine-generated files will impact on the storage space required for these data.

#### 4.1.3 Documentation

Does the respondent have accompanying documentation for their data?

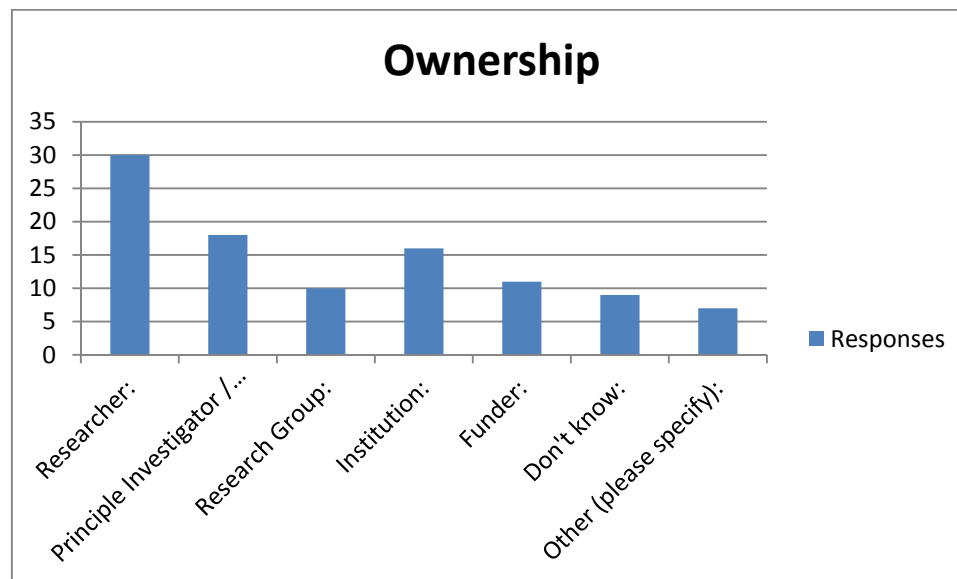
Yes:	45.5%	25
No:	47.3%	26
Don't know:	7.3%	4

This split is fairly even, although some interviewees did mention concerns over possible misuse/mis-interpretation of data as a concern. The quality of the documentation was not assessed (or self-assessed). The importance of documentation is raised again under sharing of research data, where, as noted in interviews, context surrounding the data is important.

This may be improved by using (and following) data management plans, although only three of the 22 interviewees had experience with DMPs (2 as part of a grant application, the other as demonstrating good governance when preparing the project). 69.1% of survey respondents don't have any data management plans or policies (ie. procedures for how to create, maintain, secure and preserve the data).

This is an area which is ideally suited to training, particularly for grant applicants and for PhD and Early Career Researchers.

#### 4.1.4 Perceived ownership of data



One survey respondent asked “What do you mean by “own”?” Another interviewee noted that data has always been the by-product of research, not a key output so ownership and management have often been neglected/irrelevant to the outcomes of the project. In general, if not owned by a third party, the data is considered to have shared ownership amongst its creators, often informally. Ownership may be seen as less of an issue than the provision of context surrounding the data – access by users other than the owner/s may mean the data is misconstrued or handled inappropriately when the context is not communicated.

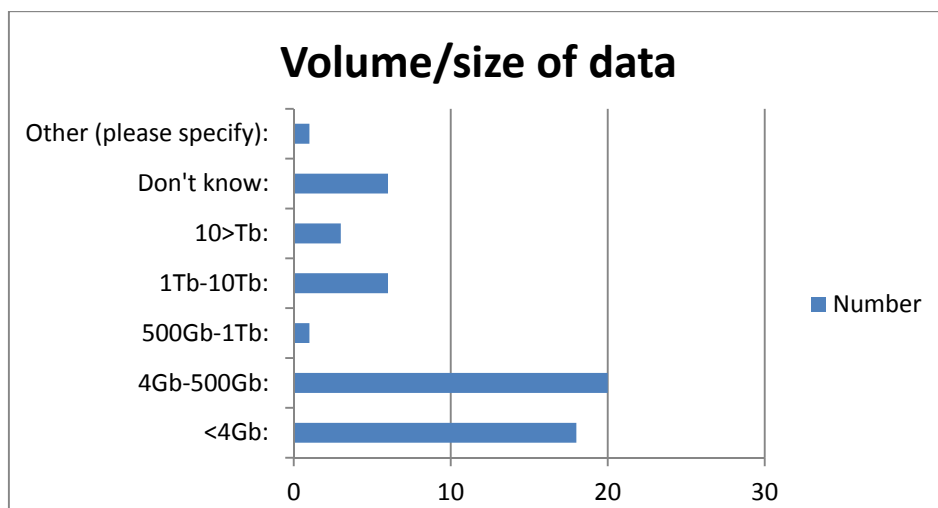
## 4.2 Data Storage and Safety.

This section gathered information on the technical activities of respondents. This included details on the size of data produced (to inform BUCS decisions), and storage and backup locations and frequency.

### 4.2.1 Volume of data

The survey asked ‘what volume of data do you estimate is currently held?’ with the sub-question of ‘Is the data expected to grow in size (ie. to next response level)?’

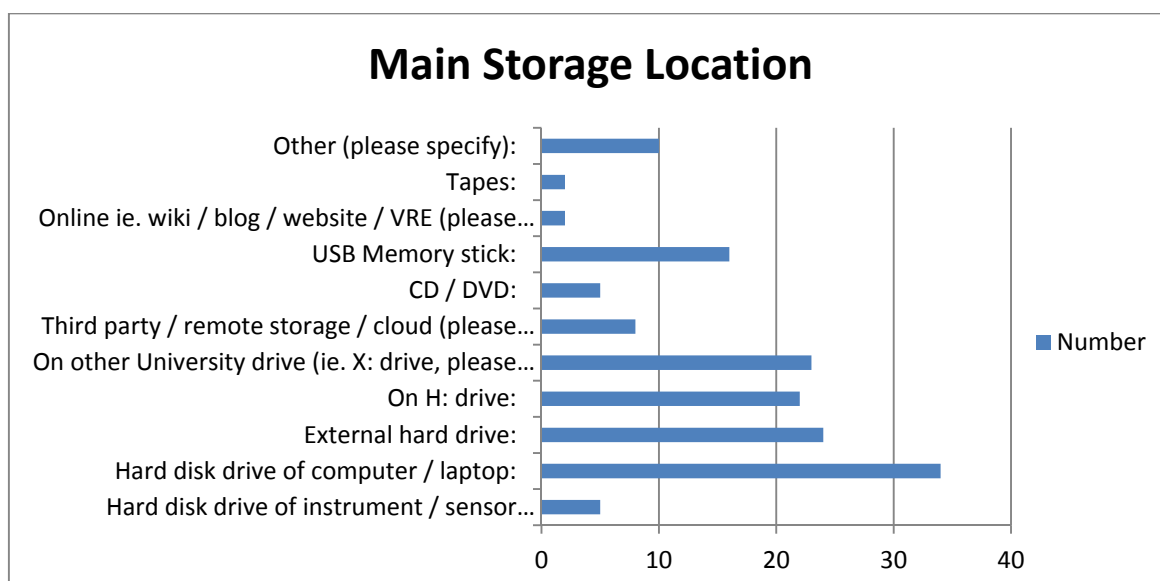




13 respondents noted they expected their files to grow in size. It is clear that the majority of respondents produce or use less than 500Gb of data. BUCS currently provide 4Gb space on personal H: drives, and up to 1Tb space on X: drives.

#### 4.2.2 Main storage location

Where are the main research data stored?



Most respondents store their research data on the hard drive of their computers, a University Drive or an external hard drive. Most people are happy with their current storage, although conversely, numerous interviewees expressed frustration at a lack of adequate storage for their research data on the H: and X: drives and cited this as a factor causing difficulty in backing up work appropriately.

There is some uncertainty about how often University drives are backed-up.

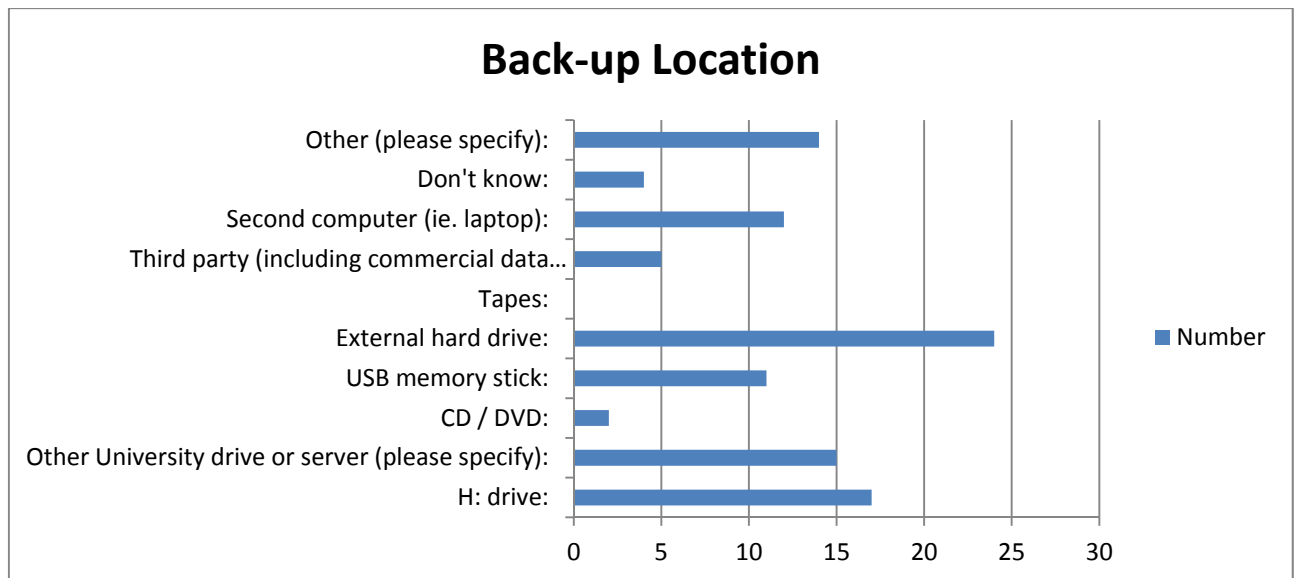
Storage on servers bought for the project, and cloud storage like Dropbox were also mentioned.

#### 4.2.3 Are you happy with your current storage?

Yes:	70.9%	39
No:	21.8%	12
Don't know:	7.3%	4

#### 4.2.4 Back-up

Where is the data backed up and how frequently does this action occur?

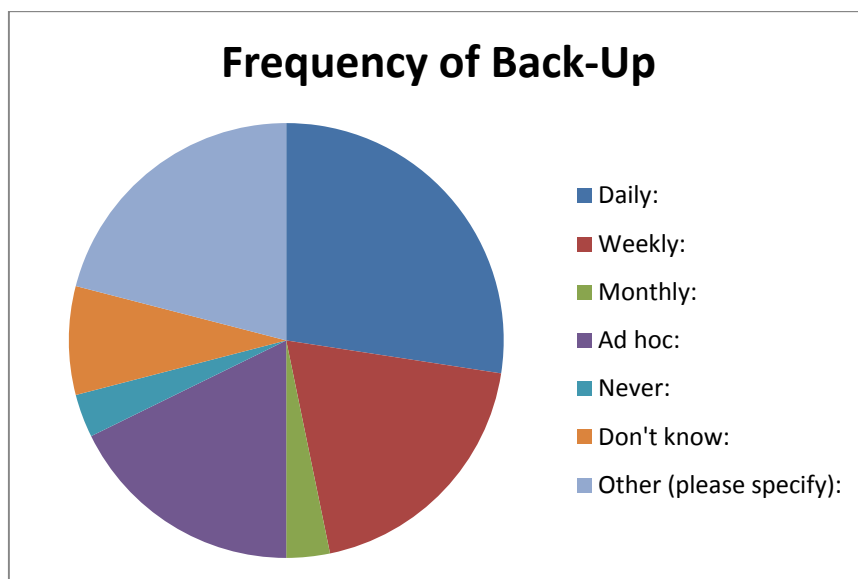


Back-up of research data on external hard drives is pervasive. Almost all interviewees backed up their data onto external/portable hard drives, with University drives as additional back-up.

Reasons for back-up choice include ease of use and portability (especially USBs). Convenience is a factor, and the ability to control or manage this space is also important.

Cloud storage locations included iDrive (.me cloud storage) and LiveDrive. Many use the H: or X: drives for back-up, but are unsure of how often or where back-up occurs.

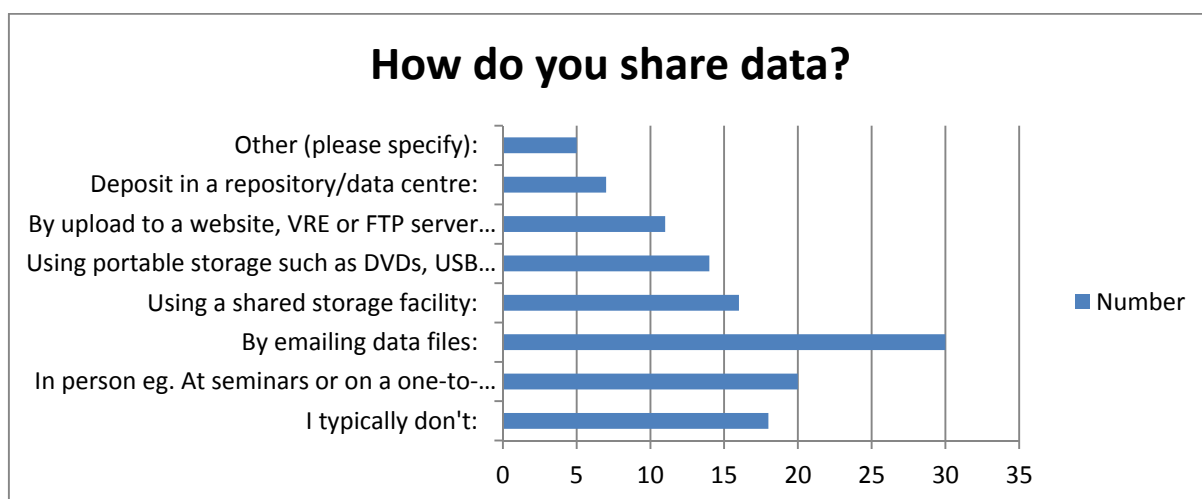
The majority of respondents back up their work daily or weekly, and many noted they rely on University processes to back up the H: and X: drives regularly. Most respondents backed-up everything each time, or backed-up their recently modified or added data only.



#### 4.3 Access and data sharing:

Research data is shared both internally (within the University) and with collaborators external to the University, and sometimes with both. A number of respondents noted the need to share data with collaborators external to the University, and noted difficulty doing this. Email was by far the most popular method. Some data creators negotiate version control methods with their collaborators, but many reported that automated version control (such as that on a wiki, or via Dropbox) is their preferred mechanism.

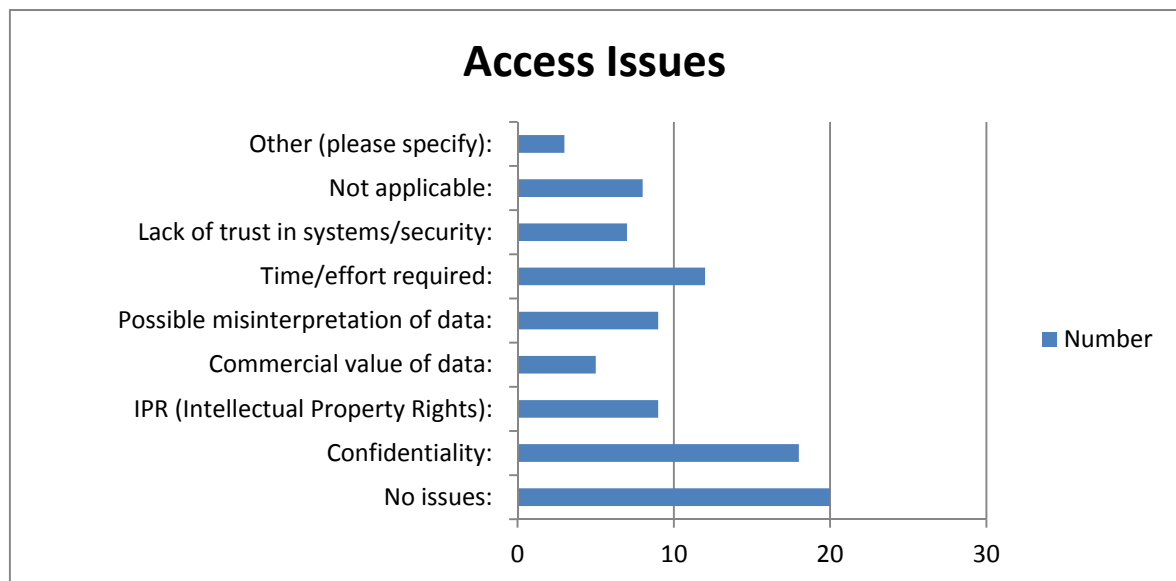
##### 4.3.1 How do you currently share your research data?



Other methods were largely based on uploading files to websites or services for online sharing.

Confidentiality was an issue highlighted by numerous interviewees so security must be addressed.

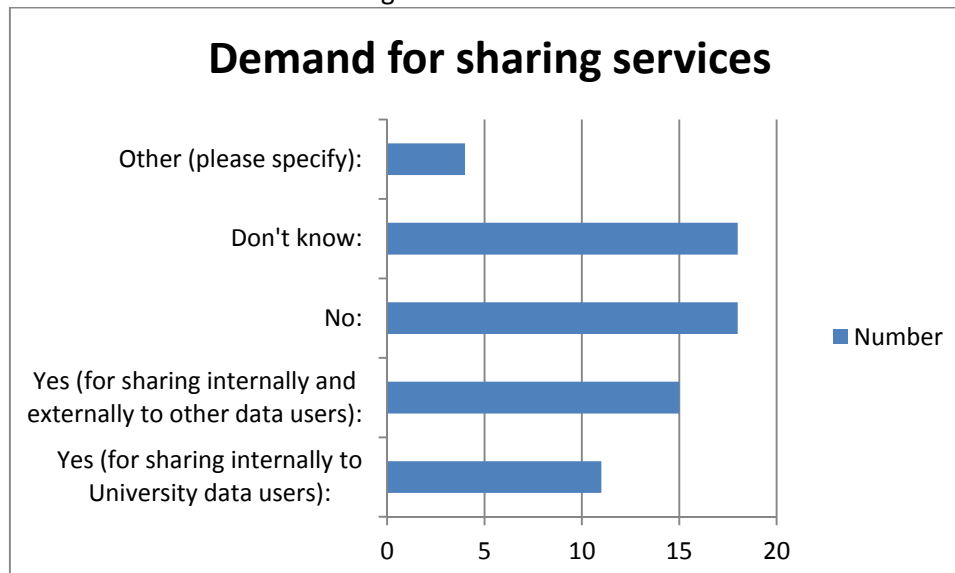
#### 4.3.2 If you share your data are there access issues of concern to you?



This response was reflected in the comments from interviewees, who largely cited confidentiality constraints as an issue. Legal requirements for the secure storage of data, and ethical concerns over anonymity were given, i.e. 'Yes, there is a confidentiality agreement with the original data sources. Interviews are initially confidential but are later anonymised and this version may not be confidential' (KLJ/1/4-2011), and 'Data from sponsors must be kept confidential and returned' (KLJ/25/3-2011).

#### 4.3.3 Is there a demand for additional services to support better data sharing or collaboration?

The next area deals with sharing services.



Those using wikis to share noted difficulties sharing with external partners.

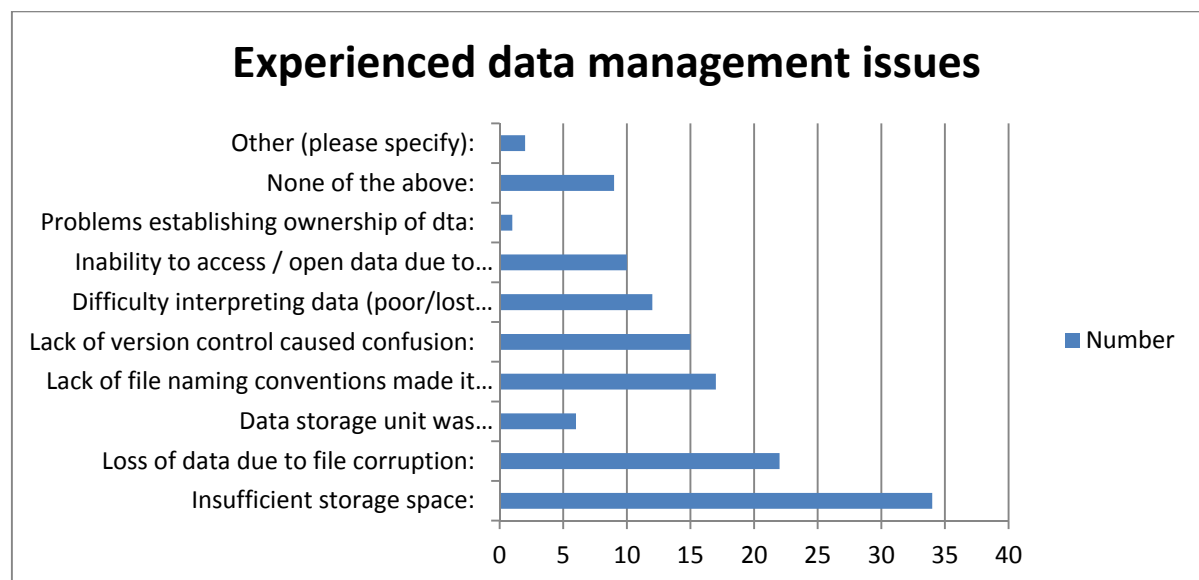
78.2% of respondents claim not to have been asked to make their electronic research data openly available outside of a publication (eg. required by a project funder). Interviewees also noted that security of data must be robust, and this is mandated for some commercial sponsors.

That said, 89.1% claim their data does need to be preserved for possible future use or access. Just over half of the respondents aren't sure or will not be making selections for keeping or disposing/deleting of data.

40% don't know how long their data should be retained for, and 48.1% claimed their data should be retained for five or more years. This is an issue that is coming forward in researcher's minds, as more report a need to consider lifespan and retention of data.

#### 4.4 Data Management

##### 4.4.1 Data issues experienced by respondents.



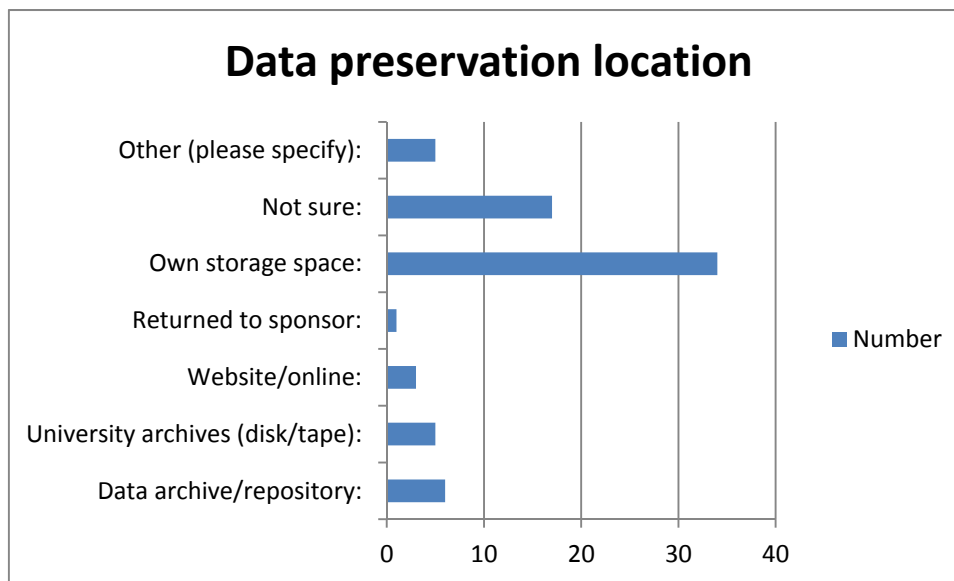
Insufficient space on the University drives, the recent email outage and account restrictions on the wiki were noted. One interviewee noted that they were currently re-producing 2 months' worth of work due to the failure of their portable storage device.

The interviews also elicited details on how sufficient users felt their research data support was. Generally this was felt to be sufficient, although the load is born by departmental IT technicians. Most issues result from hardware failure or loss.

#### 4.5 Data preservation:

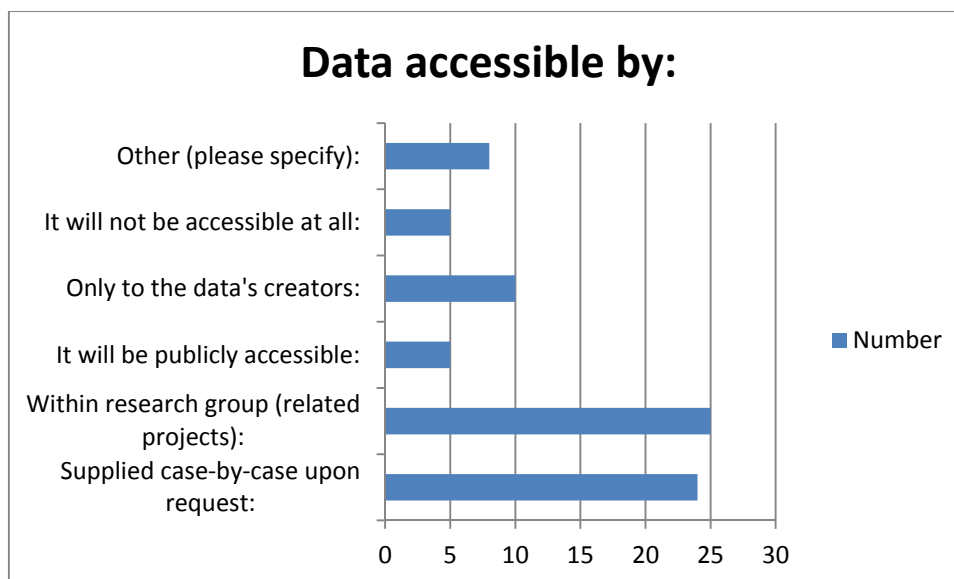
Respondents were asked where they stored (or plan to store) their data once they'd finished using it.

#### 4.5.1 Data archive or preservation location



Most respondents will be preserving their data on their own storage spaces. Combined with the lack of documentation giving the data context, and the forthcoming requirements to share and archive data for retrieval by research sponsors, this is a concern.

#### 4.5.2 In future, to whom will this data be accessible?



Most respondents currently don't see their data being used in future by anyone outside of their research group, themselves or accessible other than on a case-by-case basis.

#### 4.5.3 Preservation/Archiving of Data

Do you feel the necessary infrastructure is in place to support preservation/archiving of data?

Yes	20.0%	11
No	45.5%	25
Don't know	34.5%	19

There is clearly some uncertainty over whether the necessary infrastructure is in place. The question is quite abstract but perhaps indicates an opportunity for more specific instructions or guidelines on when, where and how data should be preserved/archived. The interviews highlighted a much stronger desire to better preserve and archive data.

Responses include 'You never know when it might be useful. University to preserve it. Readily accessible for about 5 years' (JSC/2/03-2011). Additionally 'It's use might not be immediately apparent but become useful later' or 'time series analysis may be applied at a later date' (KLJ/1/-4-2011).

## **5. Conclusions**

1. Storage space is the most often mentioned issue for research data.
2. The need to archive and share data has not been considered by most researchers – data has been a by-product of the research, usually reported via publications.
3. There is support for online storage and perhaps an institutional licence for a particular provider. Some interviewees noted that commercial products are available to all internet users that may meet researchers needs.
4. Suggested material/answers to Research Council DMPs should be available to save time/energy for grant applications.
5. Training for PhDs and early career researchers in data management is frequently requested.
6. Many respondents use BUCS drives for storing research data, and particularly for backing up working research data (in addition to external hard drives). Space availability and allocation on the X: drive aren't well understood or known about.
7. Many respondents would appreciate guidelines on how and where to archive research data so it can be easily found, retrieved and manipulated.
8. Guidance on dealing with old file formats was also requested.
9. Central licences for software (e.g. for synchronising data for back-up) would be economical.
10. BUCS departmental IT supporters are frequently the first port of call for help with issues surrounding research data.



## 6. Recommendations

1. Provision of a named person to contact for technical issues. This may be the BUCS IT supporter in a department. Support for an increasing number of Apple Mac users is also requested.
2. Whichever platform or system is used, version control of data should be automatic (ie. Dropbox). One interviewee notes "...commercial providers are available to all users, easily registered outside of university for collaborating, and work is backed-up locally and auto-synced. Why try to replicate well-functioning third party systems" (KLJ/12/04-2011).
3. Suggested material/answers to help complete Research Council DMPs should be available to guide grant applicants.
4. Training is given for PhDs and Early Career Researchers in good research data management.
5. The backup of data on University drives must be (continue to be) seamless and secure. There must be clear, concise documentation on this in order to gain the trust of researchers.
6. Archiving of data will be automatic, simple and easy.
7. Context and metadata for data should be provided to guard against mis-use or mis-interpretation.
8. Researchers would appreciate clear and open statements on University research data management policy. "I think it would be very useful to have a central policy on data management. This would give an overview at institutional level and could helpfully inform individual practice across the University. I think the other area where possible improvements could be made is that of archiving data once a project has ended. Some guidance on standard labelling of data and advice on migration of formats, etc. to facilitate long-term preservation and access would be very useful." (LR/5/04-2011)

**Project Proposal – Data Asset Framework audit**

**Aim:**

To undertake an audit of research data holdings in order to inform University policy on research data management.

**Background:**

Research Data Management deals with organising and preserving research data<sup>1</sup> so that it can be available for future use. Properly curated research data can be better shared among the wider research community, enhancing the long-term value of the work.

Initial investigations into current data practices at Bath are anecdotal but suggest that the research data is held on a variety of storage spaces with little or no planning, context or metadata. Potential issues such as funding body requirements, value for money, reputational costs, etc., provide impetus to develop this area.

**Methodology:**

The Data Asset Framework (DAF) is a JISC funded tool created for Higher Education Institutions to help them take stock of data holdings and ensure appropriate data management practices are in place. It is a useful tool to engage researchers in data curation and to scope their data management requirements.

What is DAF?

The Data Asset Framework is a set of methods to:

- find out what data assets are being created and held within institutions;
- explore how those data are stored, managed, shared and reused;
- identify any risks e.g. misuse, data loss or irretrievability;
- learn about researchers' attitudes towards data creation and sharing;
- suggest ways to improve ongoing data management.

The DAF methodology is written for information professionals. It is envisaged the person undertaking a survey would have either a qualification in library, archive or information management, or significant experience working with data. Such skills are needed to understand the information lifecycle and identify risks in existing research workflows and data management practices.

**Project requirements and costs:**

Expected timeframe for project: 8 weeks (possibly March/April 2011)

*Anticipated Staffing:*

Research Publications Librarian: project management (not costed, part of remit)

- 1 full-time researcher/interviewer for 5 weeks (potentially approximately 3 days per week over 8 weeks, or 2 researchers part-time pro-rata) Grade 3 / 4.

- Opus support team time – not costed
- Subject Librarian time – not costed

*Cost to participating departments/researchers*

- Estimated survey time – 15 minutes for each response
- Estimated interview time – approximately one hour for each interview (number of interviews not yet determined)

*Cost to not do the audit –*

- Ad hoc reporting of data assets insufficient to inform University strategy
- Future storage capability not relevant to researcher needs
- Training and planning for data management not offered/relevant
- Inadequate data management systems mean a potential risk of litigation with costs to reputation and revenue streams.

**Preparation work:**

- Request to appropriate University authority to conduct a survey
- Outline of survey and request for permission sent to appropriate Heads of Departments
- May require approval from Research Ethics committee
- Training of auditors for interview activities, including invitation to Subject Librarians to participate in interviews

Centres to be audited may be those which have previously agreed to be part of research data audits:

- Centre for Regenerative Medicine
- Institute for Sustainable Energy and the Environment (I-SEE)
- Centre for Orthopaedic Biomechanics (information may already exist as part of UKOLN's previous work on data audits – consult with Alex Ball, UKOLN)

There is also interest from the DTC for Sustainable Chemical Technologies and the Evolutionary Biology group in Biology & Biochemistry. A group or centre from the Faculty of H&SS should also be included in this project – to be announced.

**Anticipated Project Timeline:**

*Weeks One & Two –*

- Planning, defining the purpose and scope of the survey and conducting preliminary research
  - Undertake desk research on DAF audits carried out previously, noting suggestions and tips for good practice.
  - Contact HoDs and Centre Managers to outline benefits of the DAF, expected outcomes, and encourage departmental/centre support.
  - Develop survey questions (based on DAF) and train auditors

*Weeks Three to Six –*

- Identify what data assets exist and classify them (ie. vital, important or minor) to determine where to focus efforts for more in-depth analysis.
  - Roll out survey and begin initial collation of responses to highlight potential interviewees.
  - Follow up interview with targeted researchers
  - Begin to collate collected data

*Weeks Seven to Eight –*

- Pull together the information collected and provides recommendations for improving data management
  - Finalise data collection and interpretation
  - Produce draft report for circulation to Pro-VC and Associate Deans.

**Forecast results:**

Results will help inform research data management policy and services/facilities at University-wide level, led by the Pro-Vice-Chancellor (Research) and the Associate Deans for Research.

Potential solutions are expected to centre on the creation of appropriate policy on research data management with advocacy, guidance, training and documentation to support this.

Kara Jones

Research Publications Librarian

1 December 2010

## Appendix Two - Audit Form 1: Audited organisation

Organisation identifier	Bath
Organisation name	University of Bath
Organisation address	Claverton Down Bath BA2 7AY
Organisation contact details	As above
Organisation contact name	Kara Jones, Research Publications Librarian. K.L.Jones@bath.ac.uk Ph. 01225 384897
Organisation type	University
Description of the organisation	The University of Bath is a research focused university with approximately 14,000 students and 2685 staff. Our research strengths in engineering, physical sciences, management, humanities, social sciences, health and life sciences.
<b>Data management details</b>	
Strategies	<p><b>University Research Strategy</b>  <a href="http://www.bath.ac.uk/opp/policy/Research_Strategy.pdf">http://www.bath.ac.uk/opp/policy/Research_Strategy.pdf</a></p> <p><b>Good Practice Code for Research</b> (under revision): Section 7 outlines good practice for research data management  <a href="http://www.bath.ac.uk/opp/resources.bho/goodpracticeresearch2010.pdf">http://www.bath.ac.uk/opp/resources.bho/goodpracticeresearch2010.pdf</a></p> <p><b>BUCS User Filestore Policy</b> defines allocation and management, including back up and virus checking, of the general user filestores for home directories and email provided by Computing Services.:  <a href="http://www.bath.ac.uk/bucs/aboutbucs/policies/userfilestorepolicy/">http://www.bath.ac.uk/bucs/aboutbucs/policies/userfilestorepolicy/</a></p> <p><b>IT Security Policy</b>  <a href="http://www.bath.ac.uk/bucs/aboutbucs/policies-guidelines/policies-it-security.html">http://www.bath.ac.uk/bucs/aboutbucs/policies-guidelines/policies-it-security.html</a></p>
Responsibilities	Details of staff who have clear responsibility for managing the data assets TBA
Budget	Level of resources available and details on whether it is sufficient to manage data at the present time TBA
Audit details	
Dates of the audit	16 March 2011 – 8 April 2011 (information collection from surveys and interviews)
Name of the auditor	Kara Jones
Auditor contact details	K.L.Jones@bath.ac.uk Ph. 01225 384897
Date of completing the form	11 March 2011